GREP
גרף באינדיזיין ב-10 דקות
למתחילים שרוצים להגיע רחוק

<- ->

יש כלל אחד שאתם צריכים לדעת:
כל דבר שאפשר להגדיר במילים - אפשר לחפש

דוגמאות קלילות למושג הגדרה: המילה הראשונה של כל פסקה. כל מילה שניה. מילה שמתחילה עם אותיות אב ומסתיימות באותיות הם. כל מה שנמצא בין המילה מר או הרב למילים נכבד מאוד. פסקאות שחוזרות על עצמם. מילים כפולות. וכן הלאה. כי אלו דברים שאפשר להגדיר אותם במילים ולכן גם אפשר לחפש אותם.

<- ->

אז מה זה גרף?

בגרף משתמשים בתווים כלליים, המסמלים משהו אחר במקום התו הרגיל.
תו הנקודה. סוגריים [()] וכן תווים נוספים מסמלים משהו אחר ממשמעותם הרגילה.

קדימה, עכשיו יש לנו רק 9 דקות,

ואנחנו רוצים לדעת הכל.
טוב, כמעט הכל.

אלמנטים:

\	**ביטול/הפעלת תו כללי** התווים הכללים הם אלו [ ] ( ) \ ? . + ^ $ * { } זאת אומרת שאם נרצה להשתמש בהם באינדיזיין **לא** בתור תו כללי, נשים לפניהם סלש כזה \ שיבטל אותם מלהיות תו כללי.
|	**אלטרנטיבה** זה או זה. לדוגמה: יצד|חק יחפש יצדק או יצחק. דוגמה נוספת: (יעקב|משה|אהרן) יחפש יעקב או משה או אהרן.
( )	**קבוצה (בא לידי שימוש בעיקר בהחלפה)** חיפוש: (הרב)(יעקב)(יצחק)(שליט"א) והחלף עם: $1$3$2 ייתן לנו תוצאה של הרב יצחק יעקב. המילה יצחק יהיה לפני יעקב כי החלפנו את קבוצה 3 לפני 2. המילה שליט"א נמחק כי לא שמנו אותו בהחלפה. למתקדמים - ראו למטה בסוף תווים כללים דוגמה נוספת.
(?:)	**קבוצה לא ממוספרת** חיפוש: (הרב)(?:יעקב)(יצחק)(שליט"א) והחלף עם: $1$2$3 ייתן לנו תוצאה של הרב יצחק שליט"א כי המילה יעקב לא נכלל בספירת הקבוצות. הערה חשובה: סדר הכתיבה מתחילה תמיד משמאל לימין. אז גם אם אתם כותבים מימין לשמאל שימו לב למיקום.
[א-ת]	**סדר של תווים** חיפוש: יצ[א-ה]ק ייתן לנו תוצאה של יצדק אבל לא יצחק
[^ט]	**כל תו שהוא לא אות ט** חיפוש: יצ[^ד]ק ייתן לנו תוצאה של יצחק, יציק, יצמק אבל לא יצדק
[חמף]	**אחת מאותיות אלו** חיפוש: יצ[דחי]ק ייתן לנו תוצאה של יצחק, יציק, יצדק אבל לא יצמק

תווים כללים:

\t	טאב
\r	אנטר
.	**כל תו** חיפוש: יצ.ק ייתן לנו תוצאה של יצחק, יציק, יצמק, יצדק אבל לא יצק
\w	**כל תו מכמילה** מה שלא ימצא זה רווח וכדומה
\W	כל תו שהוא לא מילה
\s	**כל תו לבן** רווחים, אנטרים וכו'
\S	כל תו שהוא לא תו לבן
\d	כל ספרה
\D	כל תו שהוא לא ספרה

מיקומים:

^	תחילת פסקה
$	סוף פסקה
\b	**בגבולות המילה (בתחילת או בסוף מילה)** אפשרות נוספת \< לתחילת מילה וכן \> לסוף מילה
\B	לא בגבולות מילה
\A	בתחילת סיפור
\Z	סוף פסקה
\z	סוף סיפור
\G	מתחיל את החיפוש מאיפה שהסתיים האחרון. או איפה שהסמן עומד ברגע

במויות:

?	**התו האחרון, 0 פעמים או פעם אחת (הגבלה)** חיפוש: אי?נדיזיין ייתן לנו תוצאה של אינדיזיין, אנדיזיין. חיפוש: יום (א-ז)?? בשבוע ייתן לנו תוצאה של יום א' בשבוע, יום ב' בשבוע, יום ג' בשבוע וכו' אך גם יום בשבוע, כי התווים בסוגריים () וכן הגרש' יש להם הגבלה אפשרית ע"י הסימן שאלה.
*	**התו האחרון, 0 פעמים או יותר פעמים** חיפוש: אי"נדיזיין ייתן לנו תוצאה של אנדיזיין, אינדיזיין, איינדיזיין אייינדיזיין וכו'
+	**התו האחרון, פעם אחת או יותר פעמים** דוגמה פופולארית לחיפוש: \(.+?\) יחפש לנו כל תו החל מסוגר פותח ועד אחרי הסוגר השני. שימו לב ששמנו \ לפני הסוגריים, היות וסוגר נחשב לתו כללי, ויש לבטל אותו מלהיות בזה אם מחפשים אותו בפועל. סימון השאלה מגביל את ה+ לסוגר השני, אחרת, הנקודה, שהוא תו כללי לכל תו אפשרי, ימשיך עד סוף הפסקה. כי בברירת מחדל החיפוש הוא מקסימלי
{1,4}	**התו האחרון, לפחות פעם אחת ולא יותר מ־4 פעמים. המספרים ניתנים לשינוי כמובן.** חיפוש: מש{1,3}ה ייתן לנו תוצאה של משה, מששה, מששה, אך לא מששששה או מה
{2,}	**התו האחרון, לפחות פעמיים, אך בלי הגבלת פעמים** חיפוש: מש{2,}ה ייתן לנו תוצאה של מששה, מששה, מששששה וכו' אך לא משה
{5}	**התו האחרון, בדיוק 5 פעמים** חיפוש: מש{3}ה ייתן לנו תוצאה של מששששה בלבד
\1	**מחזיר חיפוש של קבוצה** חיפוש: (הרב)(יצחק)(בן)\2 יחפש את המילים הרב יצחק בן יצחק. דוגמה יותר מתקדמת: (הרב)(משה |יצחק)(בן)\2 יחפש את המילים הרב משה בן משה או הרב יצחק בן יצחק, אבל לא את הרב משה בן יצחק או הרב יצחק בן משה.

מיקומים וירטואלים: לא נכללים בחיפוש עצמו

(?=)	**ראה אחורה** לדוגמה: (?=הרב)יצחק יחפש את המילה יצחק רק אם הוא נמצא אחרי המילה הרב.
(?!)	**שלילי - ראה אחורה** לדוגמה: (?!הרב)יצחק יחפש את המילה יצחק רק אם הוא לא נמצא אחרי המילה הרב.
(?<=)	**ראה קדימה** לדוגמה: יצחק(?<= שליטא) יחפש את המילה יצחק רק אם הוא נמצא לפני המילה שליטא.
(?<!)	**שלילי ראה קדימה** לדוגמה: יצחק(?<! שליטא) יחפש את המילה יצחק רק אם לא נמצא אחריו את המילה שליטא.
(?>)	**קבוצה אטומית** מוחק את שאר האפשרויות ברגע שהגיע לידי תשובה ומתייחס לזה בלבד. לדוגמה: א(?>בג|ב)ג ימצא אבגג בלבד ולא אבג (אישית לא הבנתי מתי זה בא לידי שימוש. בפרט שהוא לא מתייחס באינדיזיין לאלטרנטיבה של אבג במידה ואין אבגג במסמך).

לשונית החלפה:

$0	מחזיר את כל מה שמצא בחיפוש
$1	**מחזיר את מה שמצא בקבוצה 1.** לדוגמה אם חיפשנו (יום)( )(ראשון) והחלפנו ב: $3$2$1 נקבל ראשון יום

דוגמאות קטנות מחיי היום יום:

למצוא כל כתובת אימייל במסמך: \S+@\S+\.\S+
למצוא כל מספר טלפון במסמך שמחולק עם מקף: \d{2,3}-\d{7,8}
למצוא מילה כפולה: (\w+\s)\1
למצוא פסקה כפולה: ^(.+?\r)\1

אתר מעולה ללימודי גרפ:
www.regular-expressions.info